

텍스트마이닝을 이용한 미세먼지 관련 신문기사 분석

양지연

금오공과대학교 응용수학과 부교수

An Analysis of Newspaper Articles on Fine Particle Matter Using Text Mining Techniques

Ji-Yeon Yang

Associate Professor, Dept. of Applied Mathematics, Kumoh National Institute of Technology

요 약 본 연구에서는 미세먼지에 대한 신문 기사의 시대별, 신문사별 특징을 살펴보고 있다. 이를 위하여 빅카인즈에서 1995년 이후 주요 신문사들의 관련 기사를 추출하였고 텍스트마이닝, 감성분석, 회귀분석을 활용하였다. 그 결과, 2010년 이전에는 대기오염도 측정 단어나 국내 오염원 관련 단어가 많이 등장했으나 2010년대에 들어서면서 “중국”이 큰 빈도로 나타났으며 정책적 대응, 미세먼지가 건강에 미치는 영향, 관련 제품에 대한 광고·홍보, 국내 오염원에 관한 기사까지 다양한 주제의 기사가 등장했다. 중앙일보, 한겨레, 경향신문은 상대적으로 정부의 정책이나 규제와 관련된 기사가 많은 반면, 대부분의 지역지에서는 지역 자체의 배출원 및 저감대책에 관한 기사가 많았다. 본 연구 결과는 미세먼지 관련 언론보도의 추이를 살필 수 있는 기초 자료로 활용될 수 있으리라 기대한다. 향후 포스트코로나 시대의 국내 미세먼지의 상황과 관련 기사의 트렌드를 추가적으로 비교, 검토할 수 있을 것이다.

주제어 : 미세먼지, 텍스트마이닝, 신문기사, 감성분석, 회귀분석

Abstract This study aims to examine the trend and characteristics of newspaper articles concerned with fine particle matter. Newspaper articles since 1995 collected from Bigkinds were analyzed using text mining techniques, sentiment analysis and regression analysis. Air pollution measurement and domestic pollutants appeared frequently previously, but “China” became the keyword in the 2010s along with political action, the effects on the health, AD/PR, and domestic pollutants. Korea JoongAng Daily, Hankyoreh and Kyunghyang Shinmun have had more focused on political regulations whereas most regional daily newspapers on emission sources and reduction measures at the regional level. The results of this study are expected to be used as a reference for understanding the trend of newspaper articles. Future work includes further analysis and discussion of fine particle pollution condition and news reports in the post-COVID era.

Key Words : fine particle matter, text mining, newspaper, sentiment analysis, regression analysis

*This research was supported by Kumoh National Institute of Technology(2021).

*Corresponding Author : Ji-Yeon Yang(jyang@kumoh.ac.kr)

Received September 8, 2021

Revised October 31, 2021

Accepted January 20, 2022

Published January 28, 2022

1. 서론

1.1 연구배경

대기의 질은 국민의 건강 뿐 아니라 삶의 질에 직접적인 영향을 미치기 때문에 국내 대기오염에 대한 지속적인 관심과 대응방안이 요구되고 있다. 특히 미세먼지는 쉽게 체내로 흡입되어 각종 질환을 유발하고 인체의 면역 기능을 약화시키는 것으로 보고되고 있다[1, 2].

미세먼지는 지름이 10 μm 보다 작은 PM10과 2.5 μm 보다 작은 PM2.5로 나뉘며, 주로 석탄, 석유와 같은 화석연료의 연소 과정이나 공장, 자동차 등의 배출 과정에서 많이 발생한다 [3-4]. 세계보건기구(World Health Organization, WHO)는 1987년부터 대기질 가이드라인을 제시함으로써 미세먼지의 위험성을 강조해 왔고, 2013년에는 산하의 국제암연구소(International Agency for Research on Cancer, IARC)에서 PM10과 PM2.5 모두를 1군 발암물질로 지정하였다 [4]. 실제로 [5]에서는 미국 6개 도시에서 미세먼지와 사망률간의 양의 상관관계를 밝히고 있으며, [6]에서도 미국 20개 도시에서 PM10과 사망률 간에 관련성이 있음을 기술하고 있다. 특히 PM10의 농도가 10 $\mu\text{g}/\text{m}^3$ 증가시, 전체사망률이 0.51% 증가한다고 보고한다. 또한 [7]에서는 PM2.5의 평균 농도가 10 $\mu\text{g}/\text{m}^3$ 증가할 때 폐암 발생의 비교위험도가 1.37로 증가하며, [8]에서는 위험도가 1.08-1.11로 증가함을 입증하고 있다[9].

정부에서는 미세먼지의 위험성과 경각심이 부각됨에 따라, 2013년 7월 "대기환경보전법 시행령"을 개정하였고 시범사업을 거쳐 2014년 2월부터 PM10의 예보제를, 2015년 1월부터는 PM2.5의 예보제를 실시하고 있다 [10]. 2017년 9월에는 "미세먼지 관리 종합대책"을 마련하여, 2022년까지 핵심 배출원인 발전, 산업, 수송, 생활 부문에서 미세먼지 배출량 30% 감축을 목표로 한 중장기 대책을 제시하고 있다. 2018년 1월에는 "비상·상시 미세먼지 관리 강화대책"을, 2019년 2월에는 "미세먼지 특별법"을 잇달아 발표하였으며, 2020년 2월에는 "2020년 미세먼지 총력 대응 계획"을 수립하여 미세먼지 저감 대책을 추진해오고 있다 [11]. 하지만 이러한 정책에도 불구하고, 조사 결과 국민들의 80% 이상이 미세먼지에 대한 불안감을 나타내는 것으로 보고되고 있으며, 국내에서 이루어지는 미세먼지 관련 연구 결과가 국민들에게 충분히 전달되지 않는 것으로 평가되고 있다 [12-14].

1.2 선행연구

언론은 여론 형성에 큰 영향을 미치기도 하지만 반대로 사회구성원의 의견을 수렴, 반영하여 공론화하는 역할도 한다. [15]에서는 2016년 7월까지의 자료를 이용하여 미세먼지 관련 기사의 폭발적인 증가를 보고하며, 이에 근거하여 미세먼지 이슈의 사회적 중요도를 제시하고 있다. 하지만 미세먼지와 관련해서는 언론이 여론을 올바르게 반영하기보다는 미세먼지의 위험이 언론에 의해 재구성되었다는 비판도 있다 [16, 17]. 이 연구들에 의하면 실제 대기질은 이전과 비교해 크게 변화가 없거나 오히려 좋아졌지만, 폭발적으로 증가한 언론 보도 수나 중국발 미세먼지에 지나친 주의를 끌면서 사람들의 과도한 공포심을 자극하고 있다고 주장한다. 또한 [18]에서는 2013년 1월부터 2017년 5월까지의 미세먼지 관련 보도를 대상으로 하여 언론의 보도 특징을 검토하고 있다. 여기서는 언론의 보도가 크게 미세먼지의 원인을 국내에서 찾고 그에 대한 대책을 촉구하는 "에너지정책" 프레임과 그 원인이 국외인 중국에 있음을 강조하는 "겨울중국" 프레임을 사용하는 것으로 보고하고 있다. 문제는 기사보도가 하나의 프레임만을 강조하고 있어서, 독자들이 미세먼지 이슈가 다중적인 원인을 가지고 있고 여러 부면의 대책을 필요로 한다는 점을 간과할 우려가 있다고 지적한다.

그럼에도 불구하고 여러 기존 연구들에서는 미세먼지가 우리 생활의 여러 부면에 영향을 끼치고 있다는 점을 확인해주고 있어 미세먼지의 위험성을 간과해서는 안 될 것으로 사료된다. 대기오염과 건강에 관한 연구를 보면, 미세먼지는 2012년 3백만 명 가량의 사망에 기여했다고 보고되어지고 있다[19, 20]. 미세먼지는 호흡기질환, 심혈관질환에 영향을 미칠 뿐 아니라 당뇨 등 대사성질환의 악화, 악성종양의 증가, 뇌졸중 및 인지장애와 같은 중추신경계의 이상, 미숙아 출산의 증가, 우울증과 같은 정신질환의 증가 등 인체 건강에 광범위하게 영향을 미치고 있음이 보고되고 있다 [20, 21].

대기오염은 사회적, 경제적 비용도 초래하고 있는데, [22]에서는 2060년 기준 대기오염으로 인한 노동생산성 손실은 한국 GDP를 0.3-0.4% 가량 감소시킬 것이라고 보고 있다. 대기오염으로 작물 생산량 역시 17-18% 가량 감소할 것이며, 국내 도로교통부문 대기오염의 사회적 비용은 약 700억 달러에 달할 것으로 전망하고 있다. [23]은 2015년 기준 호흡기계 기여사망자 수로 인한 피해 비용을 7,576억 원으로 추정하고 있다. 반면 [24]는 주요 대기오염물질의 저감에 따르는 편익을 추정하고 있

는데, 이산화황, 이산화질소 및 일산화탄소를 1톤 감소시킬 경우 각각 경제적 편익은 188.4천원, 102.4천원, 105.6천원이며, 아황산가스, 이산화질소, 일산화탄소의 농도를 1ppb 감소시킬 경우 각각 경제적 편익은 968.7억 원, 368.3억 원, 6.8억 원으로 추정하고 있다. [25]는 체감 생산활동 제약 정도를 고려하여 미세먼지 주의보 발령 1일당 손해 비용을 1,586억 원 가량으로 추정하고 있다. 미세먼지 주의보 발령일수를 고려할 때, 2018년 미세먼지로 인하여 약 4조 230억 원의 비용이 발생한 것으로 나타났다. 반면 [26]은 대기오염의 개선을 위해 또는 대기오염으로 인한 건강피해를 예방하기 위해 사람들이 기꺼이 지불의사가 있는 금액(Willing To Pay: WTP)을 계산하고 있는데 가구당 WTP는 연간 38,856 원이며 수도권 지역의 연간 가치는 약 2,528억 원으로 계산하고 있다.

앞으로 대기오염에 따른 불확실성은 더 커지리라 예상되며, 이러한 불확실성은 누가 얼마만큼의 비용을 부담해야 하는지, 개선을 위해 어느 정도의 자원을 어떠한 기준에 따라 투입해야 하는지 등 사회 계층집단 간 갈등과 대립을 야기시킬 수 있다 [27, 28].

일부 미세먼지 (폭넓게 환경오염) 관련 연구들은 텍스트마이닝 기법을 활용하고 있는데, [29]에서는 6개의 키워드에 대해서 주요 신문사(동아일보, 경향신문, 한겨레) 별 내용 및 논조의 차이점을 분석하고 있다. [30]에서는 2018년도 서울시의 "미세먼지 비상저감조치"의 일환인 "대중교통 무료 정책"과 "시민 참여형 차량 2부제"에 대한 기사의 댓글을 분석함으로써, 두 교통정책에 시민들의 의견을 살펴보고, 국민청원으로 올라와 있는 글들을 분석하여 시민들이 바라는 정책의 내용을 파악하고 있다. 반면 [14]는 6개의 주요 신문사 및 방송사에서 제공하는 보도 기사와 소셜미디어(포털사이트, 블로그 등)의 게시글을 활용해서 "미세먼지 관리 종합 대책" 후의 텍스트 패턴을 분석하고 있다. 소셜미디어에서는 정책 시행 이후 건강과 관련된 게시글 및 정보가 크게 증가한 반면, 언론기사에서는 저감 조치에 대한 보도에 초점이 맞추어져 있어 대중이 원하는 정보 요구도가 충족되지 않고 있음을 지적하고 있다. 미세먼지 관련 뉴스기사에 대한 토픽 모델링도 이루어졌는데, [31]에서는 2010년 1월부터 2019년 10월까지의 152,990건의 기사를 대상으로 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 알고리즘을 적용하여 80개의 토픽을 도출하고 있다. 주요 토픽으로는 미세먼지 관련 정부의 종합대책, 미세먼지 관련 상품, 취약계층 보호 정책, R&D 및 수소차를 이용한 미세

먼지 저감 등으로 나타났다.

[28]은 네이버에서 수집한 관련 온라인 언론 기사와 댓글을 대상으로 시계열 분석을 수행하고 있다. 특징적인 점은 2016년 이후 댓글의 주된 내용이 중국에 대한 비난, 정부 대책의 부족, 정책적 실효성 부족 등으로 나타났다. [32]에서는 2017년부터 2019년까지의 네이버 환경뉴스, 한국환경정책·평가연구원(KEI)의 연구보고서, 환경부 환경백서, 환경 관련 학회지로부터 텍스트 데이터를 수집하여 LDA를 적용하고 있다. 이를 통해 환경 분야의 주제별 관심도를 측정·정리하고 각 주제어간 상관관계를 검토하고 있다.

종합적으로, 미세먼지 이슈를 비롯한 환경오염 문제는 현대 사회에서 필연적으로 발생하며 이의 해결은 우리 모두의 필수적인 과제임에 틀림없다. 효과적인 대처와 예방책을 모색하기 위해, 언론을 비롯하여 사회 구성원들의 지속적인 관심과 활발한 담론이 요구된다.

1.3 연구목적

본 연구에서는 시대별, 언론사별 미세먼지에 관한 인식 패턴 및 추구하는 가치의 변화를 파악하고, 지난 5년간 대기오염 상황과 미세먼지 관련 보도 양의 관계를 검토하고자 한다.

단 일부 연구의 접근방법처럼 미세먼지 관련 정책들의 실효성을 논하기보다는 본 연구에서는 언론에서 전달하는 관련 기사들을 분석하는 데 초점을 맞춘다. 대부분의 기존 연구들이 2010년 중반부터 코로나19 이전 텍스트 자료에 근거한 결과를 제시하고 있지만, 본 연구에서는 최근의 기사 텍스트 및 대기질 자료를 포함, 활용하여 현재의 상황을 반영하자는 데 의의가 있다.

또한 기존 연구에서는 미세먼지의 위험성 및 언론보도의 특징을 살펴보기 위해, 연구자들이 직접 개입하여 기사를 분석하고 판단하는 작업을 수행하고 있다. 그 경우 연구 결과의 정확성은 높아지지만, 연구자의 가치관이 결과에 영향을 줄 수 있고 데이터 크기가 커지면 쉽게 적용하여 검증할 수 없다는 한계가 있다. 반면 본 연구에서는 텍스트마이닝 기법을 활용하여 자동으로 보도의 특징 및 통계적 유의성을 밝히고, 이를 통해 미세먼지 관련 언론 보도의 추이를 살필 수 있는 기초 자료를 마련할 수 있으리라 판단된다.

2. 자료의 수집과 연구방법

2.1 신문기사의 텍스트 자료 수집

본 연구 분석에 사용된 자료는 한국언론진흥재단의 빅카인즈(www.bigkinds.or.kr)에서 2021년 5월 24일 기준으로 "미세먼지"로 검색되는 주요 신문사의 기사들을 대상으로 한다[33]. 기간은 1995년 1월 1일 이후로 설정하였다. 상세검색 기능에서 검색어 처리는 형태소 분석을 사용하여 제목과 본문에서 미세먼지가 등장하는 뉴스들을 검색하였다.

신문사별 관련 보도 내용 및 기조의 차이를 감안하여 보수중앙지, 진보중앙지, 경제지, 지역지를 구분하여 선정하였다. 그렇게 함으로써 표본의 편향성(sampling bias)을 미리 방지하고 대표성을 확보하고자 하였다. 이에 보수중앙지, 진보중앙지, 경제지, 지역지를 구분하여 14개의 대표 신문을 선정하였다. 보수중앙지로는 조선일보, 중앙일보, 동아일보를, 진보중앙지로는 경향신문, 한겨레를, 경제지는 매일경제, 한국경제를 포함시켰다. 지역지는 권역별에 따라 강원일보(강원), 광주일보(광주, 전라), 매일신문(대구, 경북), 대전일보(대전, 충청), 부산일보(부산, 울산, 경남), 경인일보(인천, 경기), 한라일보(제주)를 선정하였다.

2.2 대기오염 자료 수집

한국공단은 전국 162개 시군의 600개의 측정망을 구축하고, 각 측정소에서 매 시간별로 대기환경기준물질 6개 항목(아황산가스, 일산화탄소, 이산화질소, 오존, PM10, PM2.5)에 대한 대기오염도를 측정하고 있다. 측정된 자료는 국가대기오염정보 관리시스템으로 수집되어 에어코리아 웹사이트를 통해 제공되고 있다 [34]. 대기오염물질 중 PM10과 PM2.5는 입자의 크기가 각각 10 μ m, 2.5 μ m 이하인 먼지를 말하며, 현재 이들에 대한 오염도인 시간당 평균 농도(μ g/m³)가 발표되고 있다. 국내 미세먼지 예보등급에 의하면, PM10의 수치가 0-30일 경우 "좋음", 31-80일 경우 "보통", 81-150일 경우 "나쁨", 151 이상일 경우 "매우나쁨"으로 분류하고 있으며, PM2.5의 수치는 0-15일 경우 "좋음", 16-35일 경우 "보통", 36-75일 경우 "나쁨", 76 이상일 경우 "매우나쁨"으로 분류하고 있다.

본 연구에서 사용하고 있는 대기오염 데이터는 에어코리아(www.airkorea.or.kr)를 통해 획득하였다. PM2.5의 측정자료는 2015년 1월부터 공개되고 있으며, 최종 확정 측정자료는 2021년 5월 24일 기준 PM10과 PM2.5 모두 2020년까지 조회가 가능하다 [34]. 에어코

리아는 측정소별로 시간별 자료를 제공하고 있는데, 본 분석에서는 서울 중구의 2015년부터 2020년까지의 시간별 PM10과 PM2.5 데이터를 먼저 일별 산술 평균값, 최댓값으로 변환한 후, 이를 이용하여 연도별 평균값, 최댓값을 계산하였다.

2.3 텍스트마이닝 및 감성분석

텍스트마이닝은 비정형 빅데이터인 텍스트를 분석하기 위한 방법으로서, 텍스트 데이터에 숨겨진 경향이나 특징을 찾아내는 과정이라 할 수 있다. 일반적으로 텍스트마이닝에는 텍스트 분류(classification), 군집화(clustering), 감성분석(sentiment analysis), 개체명 인식(NER, Named Entity Recognition), 자동 문서요약(automatic text summarization), 특징 추출(feature extraction), 토픽 모델링(topic modeling) 등이 포함된다[35].

수집된 텍스트는 먼저 코퍼스(Corpus)를 구축하는 과정이 필요한데, 이 때 코퍼스는 구조화된 텍스트들의 집합을 의미한다. 이후 노이즈를 처리하는 과정을 거치는데, 여기에는 오타, 띄어쓰기 오류 등을 수정하거나 제거하는 과정이다. 토큰화는 문장에서 토큰(token)을 나누는 과정으로 단어, 엔그램, 문구 등이 기준이 될 수 있다. 다음으로 명사, 동사, 조사 등의 품사를 분류하는 품사 태깅, 불용어 제거, 문서에서 많이 등장하지 않거나 활용되지 않는 품사에 속하는 단어들을 제거하는 과정을 거치게 된다. 그런 후, 정제된 단어들의 출현 빈도수를 측정하여 단어-문서 행렬(Term-Document Matrix)을 구축하게 된다. 이때 빈도수 대신 문서에서의 각 단어의 중요도를 활용할 수 있는데, 예를 들어 모든 문서에서 너무 자주 등장하는 단어들에 대해서 패널티를 부여하는 방법을 사용할 수 있다.

감성분석은 텍스트마이닝의 한 기법으로 텍스트에 내재되어 있는 의견, 감성 등의 주관적인 정보를 자동으로 분석하는 과정이다. 특히 사전기반 방법은 이미 구축되어 있는 외부 감성사전을 차용하는 방법으로, 감성사전에 수록되어 있는 긍정적/중립적/부정적 단어가 문서에서 얼마나 등장하는지를 파악하여 문서의 감성을 판별해내는 방법이라 할 수 있다[36].

2.4 전처리과정

본 연구를 위해 수집된 텍스트 데이터는 다음과 같은 전처리 과정을 거쳤다. 명사 기준의 단어를 추출하였으

며, 숫자 및 기호는 제거하였다. 띄어쓰기가 잘못되었거나 올바르게 단어 추출이 되지 않은 경우에는 이를 수정하는 작업을 거쳤다.

또한 "미세", "미세먼지", "초미세먼지", "먼지" 등 너무 자주 등장하는 단어들과 분석에 크게 영향을 미치는 않는 단어들, 예를 들어 "첫 번째", "월요일", "밤사이", "하루" 등의 단어들은 삭제하였다. 기타 분석에 불필요하다고 생각되는 단어들(신문사나 기자 이름) 역시 제거되었다.

2.5 연구방법

PM10, PM2.5의 현황을 검토함으로써 기존 연구에서 지적하고 있는 대기 질의 추이를 살펴본다. 또한 미세먼지 관련 뉴스 보도의 개수와 자주 사용되는 단어들을 살펴봄으로써, 시대별(특히 코로나-19 전후) 또는 신문사 종류별로 패턴의 차이가 있는지 검토한다. 감성분석과 회귀분석을 통해, 신문사별, 시대별 또는 기사 분류별 언론 보도의 어조에 어떠한 변화가 있었는지를 통계적으로 검증한다. 본 연구에서 수행된 작업 및 분석 개요는 Fig. 1과 같다.

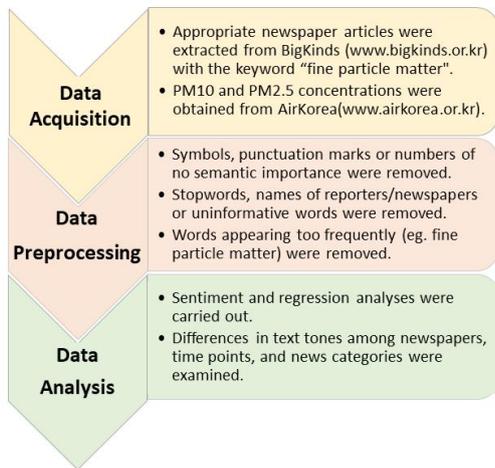


Fig. 1. Flowchart of the study

감성 사전으로는 KNU-한국어 감성사전과 easySenti R-패키지를 사용하였다. 전자는 군산대에서 개발한 사전으로 표준국어대사전의 모든 뜻풀이에 대해 Bi-LSTM을 적용하여 구축하였으며, 감성 어휘를 1-gram, 2-gram, 어구, 문형, 축약어, 이모티콘 등 다양한 형태로 제공하는 것이 특징이다 [37]. 2,597개의 "매우 긍정", 2,266개의 "긍정", 154개의 "중립", 5,029개의 "부정", 4,797개의 "

매우 부정" 단어를 포함한다. 본 연구의 감성분석에서는 KNU-한국어 감성사전을 주요 사전으로 사용하고 있으며, easySenti R-패키지는 결과를 검증하는 데 사용하고 있다. easySenti는 464개의 "긍정", 693개의 "부정" 단어를 포함하고 있다[38].

3. 연구결과

3.1 분석 텍스트 자료

빅카인즈에서 미세먼지로 검색된 1995년 1월 1일부터 2021년 5월 24일까지의 총 기사 개수는 63,185개이었다. 이중 중복, 예외 기사는 분석에서 제외하였는데, 중복 기사란 반복적으로 수집된 유사도가 높은 기사를 말하며, 예외 기사는 인사, 부고, 동정, 포도 등 유의미한 분석결과를 기대하기 어려운 내용을 담은 기사를 말한다 [33]. 또한 제목에 "브리핑"이 들어가 있는 기사들은 한 기사 내에 여러 주제의 요약된 뉴스들을 포함하고 있어서 미세먼지 관련 내용만을 추려내기 어려웠다. 그래서 이러한 기사들도 분석에서 제외하였다. 추가적으로 기사의 제목과 키워드를 검토하여 미세먼지가 기사의 주된 내용이 아닐 경우, 해당 기사도 추가적으로 분석에서 제외하였다. 최종적인 분석에 사용된 기사는 32,563개이다.

3.2 대기오염 자료

지난 5년간의 대기질 현황을 Fig. 2에서 확인할 수 있다. 현재 PM2.5의 최종확정 측정자료가 2015년부터 제공되고 있어서, 대기질 관련 분석의 시작 시점을 2015년으로 삼았다.

첫 번째 패널에서는 PM10의 연평균값(파란 점선)과 PM10이 151 $\mu\text{g}/\text{m}^3$ 이상을 기록하여 "매우나쁨"으로 분류된 연간 일수(빨간 점선)를 제시하고 있다 [34]. 이 두 기준으로 보면, 약간의 등락은 있지만 대체적으로 대기질의 상황은 2018년까지 개선되다가, 코로나19 직전인 2019년에 악화되었다. 하지만 코로나19 이후로 다시 PM10의 연평균값 및 "매우나쁨" 기록 일수가 감소세로 돌아서고 있음을 확인할 수 있다. 두 번째 패널에서는 연간 PM10의 최고기록을 나타낸다. 5년 내내 연 최고기록이 "매우나쁨"의 기준치인 151을 넘어서고는 있지만 그 값이 지속적으로 감소하고 있어 PM10을 기준으로 보면 대기질의 상황은 지난 5년간 크게 악화된 것 같아 보이는 않는다.

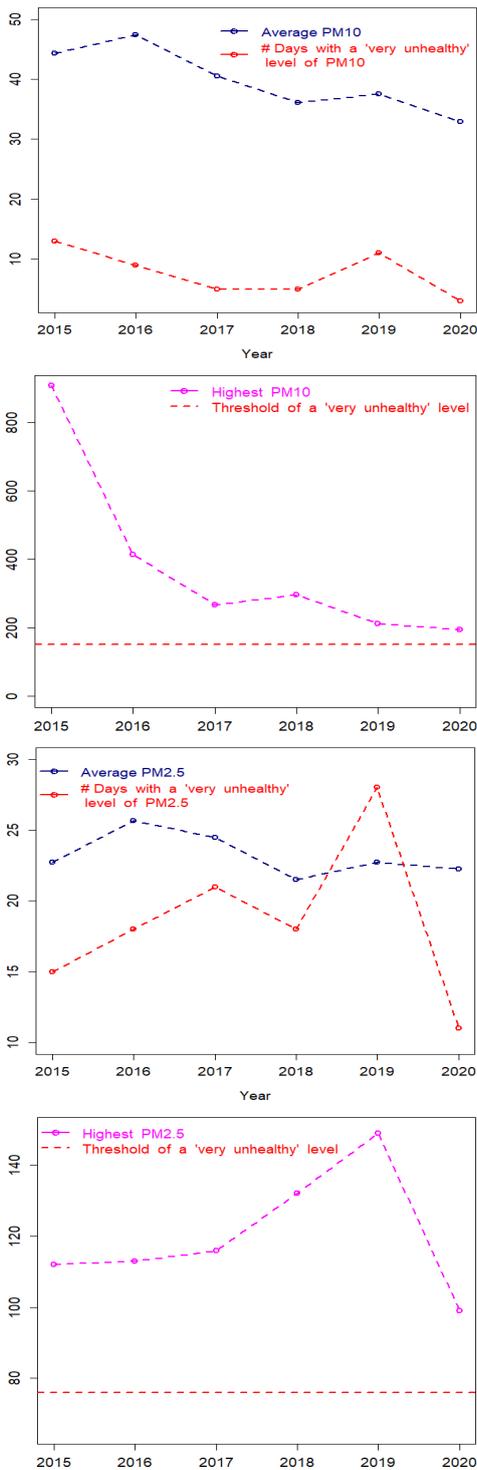


Fig. 2. Air quality over the years in Korea

하지만 PM2.5를 검토하면 이야기가 약간 달라진다. 물론 연평균값으로 보면 기존 연구 [16]에서 지적하듯이 크게 상황이 나빠진 것처럼 보이지 않는다. 하지만 PM2.5가 76 $\mu\text{g}/\text{m}^3$ 이상인 "매우나쁨"으로 분류된 일수(세 번째 패널의 빨간 점선) 및 연 최고치(네 번째 패널)를 보면, 대기질은 2018년, 2020년을 제외할 경우 전반적으로 계속 안 좋아지는 것으로 보인다. 2019년에 PM2.5의 최고치는 "매우나쁨"의 기준인 76을 훨씬 상회한 149를 기록하였다. 2020년에 상황이 좋아진 것은 코로나19로 일시적으로 나타난 현상일 수 있으며, 이는 코로나19가 종식될 경우 언제 다시 국내 PM2.5 상황이 악화될 수 있음을 시사한다. 우리의 건강에 더 심각한 영향을 미치는 것은 PM10보다 PM2.5이다[4]. 또 미세먼지 상황이 안 좋은 날이 많아질수록 야외활동에 제약이 되는 일수가 많아지는 것을 의미하므로, 위험 기준치를 상회하는 일수가 평균값보다 우리 국민들이 직접적으로 체감하는 측정치일 수 있다. 이러한 점들을 고려하면, 기존 연구들의 주장과는 달리 지난 5년간 대기 질은 결코 이전과 비교해 변화가 없거나 오히려 좋아졌다고 할 수 없을 것이다. 실제로 국민들이 체감하는 대기 오염도는 악화된 것으로 보이며, 관련 기사 수의 증가는 이러한 여론을 반영하고 공론화하는 과정이라 보여진다. 이에 미세먼지의 위험이 언론에 의해 재검토되어야 할 것으로 판단된다.

3.3 신문기사 보도 수

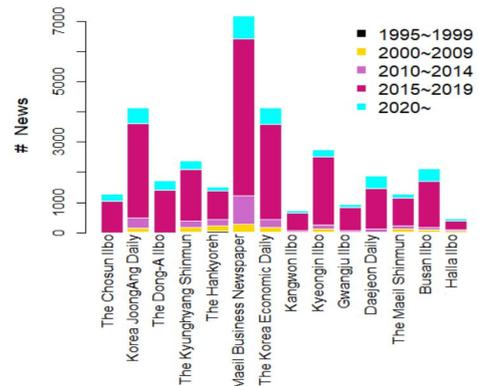


Fig. 3. The number of newspaper articles

본 분석에 사용된 신문사별 뉴스 기사의 개수를 Fig. 3에서 제시하고 있다. 시대별 추이를 살펴보기 위해, 총

분석 기간인 1995년 1월 1일부터 2021년 5월 24일까지를 5개의 시대인 1990년대, 2000년대, 2010년대 전반기(2010-2014), 2010년대 후반기(2015-2019), 2020년대로 나누고 있다. 1995년 전에는 미세먼지에 대한 보도가 거의 없어서, 그 이전을 분석 기간에 포함시키는 것이 크게 의미가 없다고 판단하여 분석 시작점을 1995년부터 설정하였다.

모든 신문사에서 1990년대, 2000년대에는 미세먼지에 대한 기사수가 많지 않다가 2010년대부터 조금씩 증가하는 추세를 보인다. 특히 2013년에 "대기환경보전법 시행령"이 개정되면서 "미세먼지 예보 시범제"가 실시되고, 중국발 스모그 현상이 사회적 이슈로 대두되면서 관련 기사수가 늘어난 것으로 보인다 [39]. 2010년대 후반에는 모든 신문사에서 관련 기사수가 급격히 증가한다. 이러한 증가 추세는, 아직 데이터가 많이 축적되지는 않았지만, 2020년대 이후 다소 꺾인 것처럼 보인다. 코로나19 이후 미세먼지 상황이 크게 개선됨에 따라 관련 기사수도 줄어든 것으로 판단된다. 신문사별로는 경제지(매일경제, 한국경제)와 중앙일보가 많은 관련 기사를 실고 있었다. 지역지에서는 경인일보와 부산일보가 많은 기사수를 보였다.

3.4 주요단어

본 분석에 사용된 뉴스 기사에서 많이 등장하는 단어들을 살펴봄으로써 시대별, 기사 분류별 특징을 찾아보고자 한다. 단순히 날씨 정보를 제공하는 6,000개의 기사는 제외하여 26,563개의 신문기사를 분석 대상으로 하고 있다.

시대별로 자주 등장하는 단어들을 살펴보면(Table 1), 1990년대에는 대기오염도 측정(오염도, ppm 등), 국내

Table 1. Top 10 words excluding institutions or regions over time (in order of frequency).

1995-1999	pollution level, vehicle, ppm, air pollution, nitrogen dioxide, permissible level, pollution, respiratory, standard, carbon monoxide
2000-2009	vehicle, respiratory, pollution level, nitrogen dioxide, air pollution, carbon monoxide, China, ppm, permissible level, atmosphere
2010-2014	China, respiratory, density, vehicle, from China, pollutant, heavy metal, the elderly, springtime, watch
2015-2019	China, vehicle, purifier, power plant, air purifier, respiratory, watch, emissions, consumer, diesel car
2020-	COVID19, China, vehicle, purifier, emissions, consumer, respiratory, citizens, workplace

오염원(자동차, 이산화질소, 일산화탄소 등)과 관련된 단어가 많이 등장하고 있다. 2000년 초반 역시 비슷하지만, 이때부터 주요 단어에 중국이라는 단어가 등장하고 호흡기가 빈도순위 상위를 기록하고 있다.

반면 2010년대에 들어와서는 중국이 가장 많은 빈도로 관련 기사에 등장하고 있다. 2010년대 후반기에는 중국이라는 단어가 2,800번 이상 등장한다. 호흡기, 주의보 등의 등장으로 보아 미세먼지가 건강에 미치는 영향에 관한 기사도 많이 늘어났으며 청정기, 경유차, 소비자, 청소기 등 환경 관련 제품에 대한 단어가 많이 등장하고 있어 광고·홍보형 기사가 크게 늘었음을 확인할 수 있다. 그 외 국내 오염원(자동차, 발전소 등)에 관한 기사도 꾸준히 있었음을 알 수 있다.

2020년 이후로 코로나19 직전보다 관련 기사수가 크게 줄었지만, 여전히 중국이라는 단어가 상위를 차지하고 있다. 국내 오염원(자동차, 배출량, 사업장), 환경 관련 제품(청정기, 소비자)에 대한 기사도 여전히 나타나고 있다. 추가적으로 기사에서 자주 등장하는 조직 및 지역 관련 상위 단어를 살펴보았는데, 1990년대에는 환경부, 서울, 서울시, 대구, 부산 등, 2000년대에는 환경부, 서울, 정부, 기상청, 서울시 등이 많이 나타났다. 2010년대 초반에는 환경부, 기상청, 정부, 국립환경과학원, 중국 등, 2010년대 후반에는 정부, 환경부, 중국, 국회, 서울시 등이 자주 등장하였고, 2020년 이후로는 정부, 환경부, 중국, 서울, 한국 등의 단어가 상위 빈도를 보였다. 1990년대, 2000년대에는 환경부, 기상청, 지방자치단체가 주로 언급된 반면, 2010년대부터 중국이 등장하고 정부, 국회가 더 자주 등장함에 따라 미세먼지가 정치적, 국제적 이슈로 부각되었음을 확인할 수 있다.

빅카인즈에서 제공하는 "통합분류1"에 따라 미세먼지 관련 기사를 분류할 경우, 각 분류에 속하는 기사들에서 많이 등장하는 단어들을 Fig. 4에서 확인할 수 있다. 거의 모든 분류에서 중국이라는 단어가 등장한다. 그중 "정치" 분류에서는 문재인 대통령, 한국당을 비롯하여 발전소, 배출량, 특별법 등 단어의 등장으로 보아 정치적 개입이나 정책적 규제에 관련한 내용이 주를 이룬다. "경제" 분류는 청정기, 자동차, 청소기, 전기차, 화장품 등 호흡기 및 피부 관련 제품에 대한 광고·홍보 내용이 많은 것으로 파악된다. "사회" 분류는 순수하게 날씨 정보만을 제공하는 기사들을 제외했음에도 불구하고, 미세먼지 관련한 많은 기사가 이 카테고리에 속하였다. 대기오염 상태 및 호흡기질환에 대한 주의보, 노후차량과 발전소, 사업장 배출 허용기준 등의 단속규정, 비상저감조치 관련

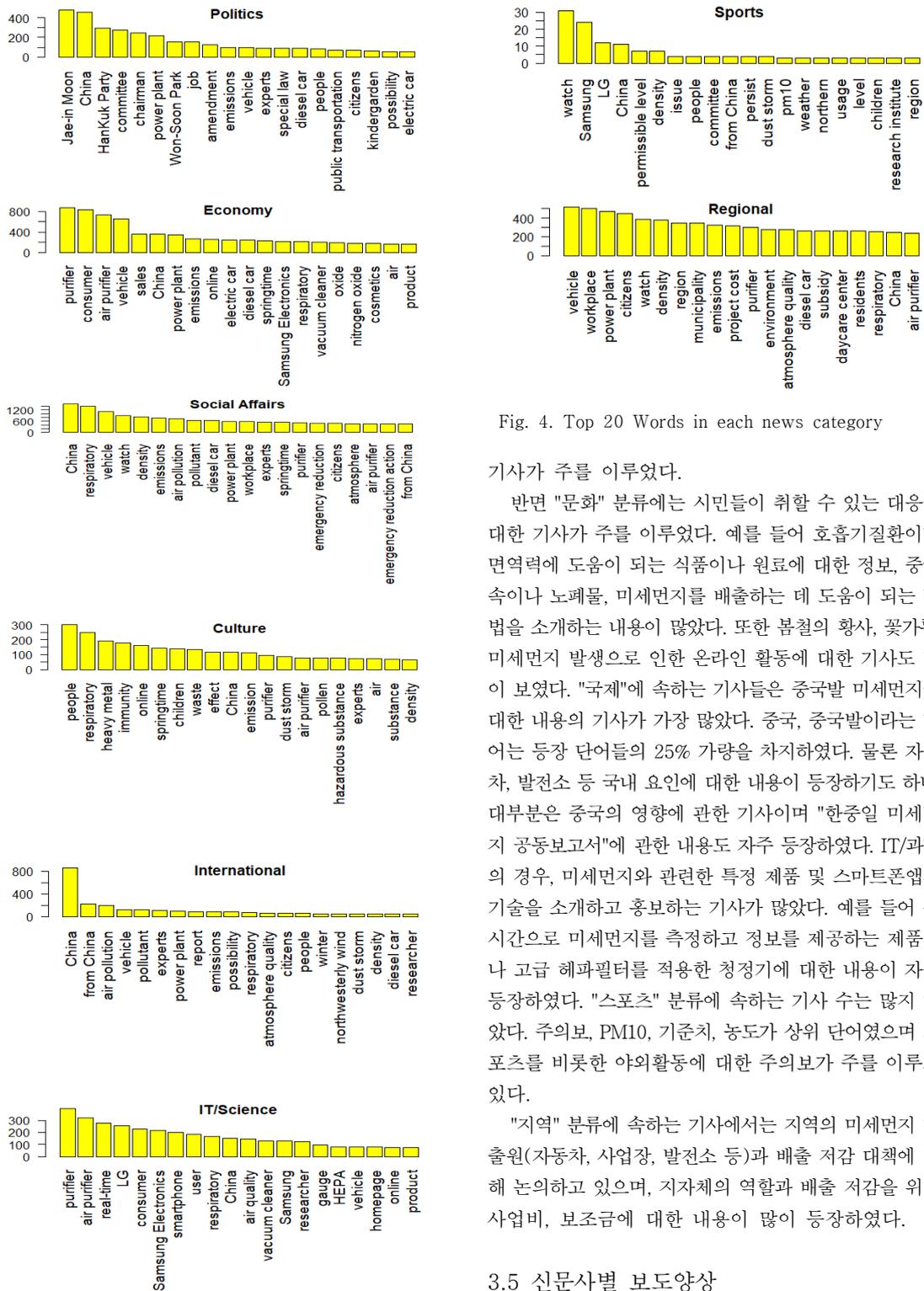


Fig. 4. Top 20 Words in each news category

기사가 주를 이루었다.

반면 "문화" 분류에는 시민들이 취할 수 있는 대응에 대한 기사가 주를 이루었다. 예를 들어 호흡기질환이나 면역력에 도움이 되는 식품이나 원료에 대한 정보, 중금속이나 노폐물, 미세먼지를 배출하는 데 도움이 되는 방법을 소개하는 내용이 많았다. 또한 봄철의 황사, 꽃가루, 미세먼지 발생으로 인한 온라인 활동에 대한 기사도 많이 보였다. "국제"에 속하는 기사들은 중국발 미세먼지에 대한 내용의 기사가 가장 많았다. 중국, 중국발이라는 단어는 등장 단어들의 25% 가량을 차지하였다. 물론 자동차, 발전소 등 국내 요인에 대한 내용이 등장하기도 하나, 대부분은 중국의 영향에 관한 기사이며 "한중일 미세먼지 공동보고서"에 관한 내용도 자주 등장하였다. IT/과학의 경우, 미세먼지와 관련한 특정 제품 및 스마트폰앱의 기술을 소개하고 홍보하는 기사가 많았다. 예를 들어 실시간으로 미세먼지를 측정하고 정보를 제공하는 제품이나 고급 헤파필터를 적용한 청정기에 대한 내용이 자주 등장하였다. "스포츠" 분류에 속하는 기사 수는 많지 않았다. 주의보, PM10, 기준치, 농도가 상위 단어였으며 스포츠를 비롯한 야외활동에 대한 주의보가 주를 이루고 있다.

"지역" 분류에 속하는 기사에서는 지역의 미세먼지 배출원(자동차, 사업장, 발전소 등)과 배출 저감 대책에 대해 논의하고 있으며, 지자체의 역할과 배출 저감을 위한 사업비, 보조금에 대한 내용이 많이 등장하였다.

3.5 신문사별 보도양상

미세먼지 관련 기사의 통합분류의 분포를 살펴봄으로

써, 신문사별로 어떠한 기사가 많이 보도되었는지 그 특징을 살펴보고자 한다. 관련 기사가 많이 나온 두 기간인 2015년-2019년, 2020년-을 대상으로 삼았다. 그렇게 함으로써 코로나19의 영향까지 확인할 수 있다고 판단하였다. 단순한 날씨 기사는 이번에도 분석에서 제외하였다.

Table 2는 2015년-2019년 동안 신문사별 기사의 통합분류 비율을 보여준다. 지면 제약상 표에서 기사의 통합분류는 축약해서 기술하고 있는데, 각각 Pol: 정치, Econ: 경제, Soc: 사회, Cul: 문화, Int: 국제, IT/Sci: IT과학, Spo: 스포츠, Reg: 지역을 나타낸다.

표에서 보듯이, 신문사별로 편차가 있기는 하지만 전반적으로 중앙지와 경제지의 경우 관련 기사가 "사회", "경제", "정치", "지역" 순으로 나타나는 것으로 보인다. 반면에 지역지는 "지역", "사회"에 속하는 기사들을 많이

보도하였다.

각 분류별로 살펴볼 경우, "정치" 관련 기사를 많이 보도한 신문사는 조선일보, 한겨레이며, 경제지와 지역지는 "정치" 관련 기사의 비율이 상대적으로 낮았다. "경제" 분류에서는 한국경제, 부산일보, 조선일보, 동아일보, 매일경제 순으로 기사들을 많이 냈으며, 상대적으로 중앙지에서 한겨레와 경향신문, 지역지(단 부산일보를 제외)가 "경제" 비율이 낮았다.

"사회" 분류의 경우, 중앙일보, 한겨레, 경향신문이 관련 기사를 많이 보도하였다. 반면 "국제" 분류는 전반적으로 중앙지에서 비율이 높으며, 상대적으로 경제지와 지역지가 그 비율이 낮은 것으로 나타났다. "IT/과학" 분류는 조선일보, 매일경제, 한국경제, 동아일보 순으로 보도 비율이 높았다. "지역" 분류는 지역지에서 비율이 높았고 중앙일보가 상대적으로 관련 보도 비율이 낮은 것으로 확인된다.

Table 2. The proportions of news categories by newspaper (2015-2019)

(%)	Pol	Econ	Soc	Cul
ChosunIlbo	13.3	25.2	26.2	3.5
JoongAng Daily	10.3	13.5	45.5	9.8
Dong-A Ilbo	6.7	24.7	36.1	6.9
Kyunghyang Shinmun	9.7	17.8	38.4	5.5
Hankyoreh	13.0	11.3	43.0	8.7
Maeil Business	8.0	24.1	32.9	9.9
Korea Economic Daily	7.0	30.3	28.5	5.5
Kangwon Ilbo	6.2	6.6	31.2	6.2
Kyeongin Ilbo	8.5	8.6	30.5	4.6
Gwangju Ilbo	5.3	13.5	19.0	6.1
Daejeon Daily	6.4	8.9	20.2	3.6
The Maeil Shinmun	7.7	11.5	27.1	8.6
Busan Ilbo	4.8	24.8	22.2	6.9
Halla Ilbo	3.2	6.8	27.4	4.6
	Int	IT/Sci	Spo	Reg
ChosunIlbo	6.3	10.2	0.4	12.3
JoongAng Daily	6.9	5.7	0.7	5.8
Dong-A Ilbo	4.2	8.1	0.9	11.3
Kyunghyang Shinmun	6.0	6.3	0.5	14.7
Hankyoreh	4.3	3.5	0.5	15.0
Maeil Business	4.0	8.5	0.4	10.7
Korea Economic Daily	4.4	9.4	0.6	11.8
Kangwon Ilbo	2.9	2.5	0.4	43.1
Kyeongin Ilbo	2.5	3.0	0.6	40.7
Gwangju Ilbo	0.0	3.1	1.2	51.5
Daejeon Daily	1.2	3.6	0.5	54.8
The Maeil Shinmun	2.1	2.2	1.3	39.3
Busan Ilbo	1.1	7.2	0.4	31.4
Halla Ilbo	2.1	3.9	0.7	47.0

Fig. 4에서 제시하는 분류별 기사들의 주요 단어들을 바탕으로 신문사별 보도 특징을 종합적으로 살펴보면, 중앙일보, 한겨레, 경향신문은 "정치"와 "사회" 분류 비율이 높고 반면에 "경제" 비율이 낮은 것으로 나타나, 정책적 규제나 조치 관련 내용의 기사가 많고 상대적으로 광고·홍보성 기사가 적은 것으로 보인다. 반면에 동아일보, 경제지, 부산일보의 경우 "사회"나 "문화" 분류에 속하는 기사보다는 "경제"와 "IT/과학"의 기사가 많은 것으로 나타나, 관련 최신기술 및 제품을 소개하고 홍보하는 기사가 많은 것으로 보인다. 지역지는 다소 편차가 있지만 부산일보를 제외한 대부분의 지역지에서는, 홍보성 기사가 상대적으로 적고 미세면지를 정치적 이슈로 검토하기 보다는 지역 자체의 배출원 및 저감대책에 대한 논의가 많은 것으로 파악된다.

Table 3은 2020년 이후의 기사 분포를 보여주는데, "정치"와 "사회"의 비율이 줄고, "지역"과 "IT/과학" 기사 비율이 늘어난 것으로 확인된다. 코로나19 이후 미세면지를 외부 요인(중국) 보다는 국내 요인에 좀더 초점을 맞추고 정책적 규제를 논하고 있는 것으로 보인다. 또한 관련 기술의 향상에 대한 소개 기사도 많이 증가하였다.

3.6 김성점수를 바탕으로 한 회귀분석

감성분석을 통해 미세면지 관련 기사에 나타나는 어조를 파악하고자 한다. KNU-한국어 감성사전에는 다섯 단계의 긍·부정을 나타내는 단어들을 포함하고 있는데, "매우 긍정"일 경우 +2, "긍정"에는 +1, "중립"에는 0, "부정"에는 -1, "매우 부정"에는 -2의 값을 부여하고 있다

Table 3. The proportions of news categories by newspaper (2020-)

(%)	Pol	Econ	Soc	Cul
ChosunIlbo	5.5	19.4	27.3	0.6
JoongAng Daily	4.7	19.5	31.7	8.1
Dong-A Ilbo	2.7	24.2	30.3	8.1
Kyunghyang Shinmun	3.6	15.4	29.2	6.7
Hankyoreh	6.2	10.7	32.1	14.3
Maeil Business	1.9	34.7	21.2	8.7
Korea Economic Daily	3.6	28.7	25.6	4.4
Kangwon Ilbo	2.3	4.5	23.9	6.8
Kyeongin Ilbo	2.4	5.3	27.5	3.9
Gwangju Ilbo	0.0	9.1	21.8	5.5
Daejeon Daily	1.2	8.0	18.1	2.6
The Maeil Shinmun	3.9	8.7	19.7	7.1
Busan Ilbo	0.2	22.3	20.3	1.5
Halla Ilbo	2.8	5.6	20.8	4.2
	Int	IT/Sci	Spo	Reg
ChosunIlbo	3.6	21.2	0.0	19.4
JoongAng Daily	7.8	12.8	1.5	12.2
Dong-A Ilbo	4.4	12.8	0.0	16.5
Kyunghyang Shinmun	4.3	18.2	0.4	21.7
Hankyoreh	7.1	5.4	0.0	21.4
Maeil Business	3.2	14.3	0.0	15.2
Korea Economic Daily	1.6	15.8	0.0	18.0
Kangwon Ilbo	1.1	4.5	0.0	54.5
Kyeongin Ilbo	1.9	4.3	0.0	51.2
Gwangju Ilbo	0.0	4.5	4.5	53.6
Daejeon Daily	0.2	4.7	0.0	65.0
The Maeil Shinmun	0.0	3.1	0.0	57.5
Busan Ilbo	0.7	9.3	0.7	43.6
Halla Ilbo	0.0	1.4	2.8	62.5

[37]. 각 기사에서 이러한 단어들이 나타날 때마다 해당하는 값들을 더한 합계를 감성점수로 정의하게 된다. 따라서 감성점수가 양의 값이면서 절대값이 클수록 긍정적인 어조가 강하며, 반대로 음의 값이면서 절대값이 클수록 부정적인 어조가 강하다 할 수 있다. 0에 가까우면 중립적이라 볼 수 있다.

미세먼지 기사가 많은 두 기간인 2015년-2019년과 2020년-을 분석 대상으로 삼았으며, 단순히 날씨를 전하는 기사는 제외하였다. 감성점수를 반응변수로 삼고 신문사, 기간, 기사의 통합분류를 설명변수로 한 회귀분석 결과를 Table 4에 제시하고 있다. 각 터미변수에서 중앙일보, 2020년 이후, "IT/과학"을 기준그룹으로 처리하였다. 계수 추정치 옆에 표시된 **와 *는 각각 유의수준 5%, 10% 하에서의 통계적 유의성을 나타낸다.

신문사에 해당하는 계수는 대부분 양수로 나타났는데, 이는 대부분의 신문사가 기준그룹인 중앙일보 보다 큰 감성점수를 보였음을 의미한다. 특히 모든 중앙지 및 지역지에서 통계적 유의성을 나타냈다. 반면 강원일보는 중앙일보 보다 작은 감성점수를 보였지만 통계적으로 유의하지는 않았다. 이러한 점은 중앙일보의 관련 기사가 다른 어떤 신문사들의 기사보다 부정적인 어조가 강했음을 시사한다. 반면 2015년-2019년 기간에는 2020년 이후에 비해 기사의 부정적인 어조가 강했다. 코로나19 이후 미세먼지 상황이 개선되면서 부정적인 어조도 많이 완화된 것으로 보인다.

"IT/과학" 분류에 속하는 기사들과 비교하여 "경제" 및 "지역"에 속하는 기사들은 긍정적인 어조였으며 5% 수준에서 통계적으로 유의하였다. 반면 "사회", "국제", "문화"에 속하는 기사들은 상대적으로 부정적인 어조가 강했으며, 역시 5% 수준에서 유의했다. "정치" 분류 역시 부정적인 어조가 강했는데, 10% 수준에서 통계적 유의성을 나타냈다. 감성사전으로 easySenti R-패키지를 사용할 경우에도 비슷한 결과를 얻을 수 있었는데, 지면 판계상 제시는 생략하고 있다.

Table 4. Regression analysis results.

		Estimates
Newspaper	Intercept	0.672**
	The Chosun Ilbo	1.019**
	The Dong-A Ilbo	0.894**
	The Kyunghyang Shinmun	0.780**
	The Hankyoreh	0.838**
	Maeil Business Newspaper	0.690**
	The Korea Economic Daily	0.931**
	Kangwon Ilbo	-0.065
	Kyeongin Ilbo	0.840**
	Gwangju Ilbo	0.151
	Daejeon Daily	0.697**
	The Maeil Shinmun	0.236
	Busan Ilbo	1.439**
	Halla Ilbo	0.416
Period	2015-2019	-0.515**
Category	Politics	-0.366*
	Economy	1.987**
	Social Affairs	-2.712**
	Culture	-0.517**
	International	-2.823**
	Sports	-0.244
	Regional	0.609**
	Unclassified	1.152**

4. 결론

본 연구에서는 국민의 건강과 삶의 질에 직접적인 영향을 미치는 "미세먼지" 이슈에 대한 신문 기사의 시대적, 신문사별 특성을 살펴보고 있다. 분석 대상인 모든 신문사에서 1990년대, 2000년대에 관련 기사가 많지 않다가, 2010년대에 들어 미세먼지가 사회적 이슈로 대두되면서 관련 기사수가 크게 증가했다. 하지만 코로나19 이후 미세먼지의 상황이 좋아지면서 기사수도 줄어든 상황이다. 이전에는 오염도, ppm 등의 대기오염도 측정 단위나 자동차, 이산화질소, 일산화탄소 등의 국내 오염원 관련 단어가 많이 등장하였으나, 2010년대에 들어서면서 중국이 큰 빈도로 나타나는 것으로 확인되었다. 또한 2010년대에는 중국 관련 기사 뿐 아니라 미세먼지가 건강에 미치는 영향에 관한 기사와 청정기, 경유차, 청소기 등 관련 제품에 대한 광고·홍보형 기사, 자동차, 발전소, 사업체 등 국내 오염원에 관한 기사까지 다양한 주제의 기사가 등장하여, 미세먼지가 일상생활 곳곳에 영향을 미치고 있음을 확인할 수 있었다.

신문사별로도 기사 유형의 차이가 있었다. 중앙일보, 한겨레, 경향신문은 "정치"와 "사회" 분류에 속하는 기사 비율이 높고 "경제" 비율이 낮았는데, 정부의 정책이나 규제와 관련된 기사가 많은 반면 광고·홍보성 기사가 적은 것으로 보인다. 경제지, 동아일보, 부산일보는 "경제"와 "IT/과학" 분류에 속하는 기사 비율이 높아, 상대적으로 관련 최신기술 및 제품을 소개하고 홍보하는 기사가 많은 것으로 보인다. 부산일보를 제외한 대부분의 지역지에서는 홍보성 기사 및 정치적 성향의 기사 보다 지역 자체의 배출원 및 저감대책에 관한 기사가 많았다.

코로나19 이후로는 "정치"와 "사회"의 비율이 줄고, "지역"과 "IT/과학" 기사 비율이 늘어나, 미세먼지를 국제 정치적 이슈로 다루기 보다 좀더 국내 요인에 초점을 맞춘 정책적 규제를 논하고 있는 것으로 보인다. 또한 관련 기술의 향상에 대한 소개 기사도 많이 증가하였다.

감성분석 및 회귀분석을 통해 중앙일보가 상대적으로 다른 신문사들보다 부정적인 어조를 나타냈음을 확인할 수 있었다. 또한 경제나 지역 관련 기사들 보다 사회, 국제, 문화, 정치 관련 기사들에서 언급되는 미세먼지 이슈의 부정적인 어조가 더 강했다. 하지만 코로나19 이후 기사에서 부정적인 어조가 많이 완화되었음을 확인하였다.

최근 5년간 고농도 미세먼지 발생일수가 증가하면서 국민들이 체감하는 대기 오염도는 악화된 것으로 보인다. 미세먼지 관련 기사 수의 증가는 이러한 여론을 반영하

고 공론화 과정이라 보여진다. 단 본 연구의 분석은 텍스트마이닝과 관련 기사 수의 추이에 바탕을 두고 있기 때문에, 미세먼지의 위험이 언론에 재구성되었다는 기존의 주장을 엄밀히 검증하는 데 한계가 있다. 기사 내용에 대한 구체적인 검토와 다양한 대기오염도 자료를 이용한 회귀분석, 상관분석을 통한 추가연구가 필요한 것으로 판단된다.

또한 기존 연구에서 지적하듯이 홍보, 광고성 기사나 단순히 날씨를 전달하는 기사보다는, 활발하게 문제 제기를 하고 해결책을 모색하는 기사를 많이 보도함으로써 언론의 사회적 책임을 보다 적극적으로 수행할 수 있을 것이다. 또한 미세먼지 문제는 다중적인 원인을 가지고 있고 여러 부면의 대책을 필요로 하기 때문에, 언론은 국외나 국내의 발생요인에 지나치게 초점을 맞추기 보다는 종합적인 관점에서 미세먼지 이슈를 접근할 필요가 있어 보인다. 정부 역시 언론을 통해 관련 정책 및 연구 결과를 국민들에게 충분히 전달하고 공유할 수 있도록 해야 할 것이다.

많은 관련 기존 연구들과 달리, 본 연구에서는 최근의 기사 텍스트 및 대기질 자료를 활용하여 현재의 상황이 반영된 시대별, 신문사별 언론보도 특징을 고찰하고 있으며, 연구자의 개입을 통한 기사의 분석, 분류가 아니라 텍스트마이닝 기법을 활용하여 자동으로 보도의 특징 및 통계적 유의성을 밝히고 있다는 데 의의가 있다.

정부는 정책 수요 분석을 함에 있어서 언론보도 자료를 적극적으로 활용함으로써 현재뿐 아니라 미래의 트렌드를 파악할 필요가 있다. 본 연구에서 사용한 자료는 신문사들의 관점을 반영한다고 볼 수 있으며, 본 연구의 결과를 통해 미세먼지 이슈와 관련하여 기사 카테고리 내에서의 비중은 어떠한지, 최근에 가장 관심을 받는 주요 키워드는 무엇인지, 키워드간의 빈도수 차이는 어떠한지 등에 대한 답을 얻을 수 있으리라 기대한다.

REFERENCES

- [1] National Institute of Environmental Research. (2006). *Study on the Development of Comprehensive Assessment Techniques for Air Pollution*. Republic of Korea: Ministry of Environment.
- [2] J. K. Choi, et al. (2020). Harmfulness of Particulate Matter in Disease Progression. *Journal of Life Science*, 30(2), 191-201.
DOI : 10.5352/JLS.2020.30.2.191

- [3] Republic of Korea: Ministry of Environment. (2020). *Special Act on the Reduction and Management of Fine Dust*.
- [4] Republic of Korea: Ministry of Environment. (2016). *You can see it if you know it right away. Fine Dust*.
- [5] D. W. Dockery, et al. (1993). An association between air pollution and mortality in six US cities. *New England Journal of Medicine*, 329(24), 1753–1759. DOI : 10.1056/NEJM199312093292401
- [6] J. M. Samet, et al. (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *New England Journal of Medicine*, 343(24), 1742–1749, DOI : 10.1056/NEJM200012143432401
- [7] J. Lepeule, et al. (2012). Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environmental Health Perspectives*, 120(7), 965–970, DOI : 10.1289/ehp.1104660
- [8] D. Krewski, et al. (2009). *Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality*. Research Reports: Health Effects Institute.
- [9] J. P. Myong. (2016). Health Effects of Particulate Matter. *The Korean Journal of Medicine*, 91(2), 106–113. DOI : 10.3904/kjm.2016.91.2.106
- [10] Republic of Korea: Ministry of Environment. (2013). <https://bit.ly/2VqVLUX>
- [11] Republic of Korea: Ministry of Culture, Sports and Tourism. (2020). *Korean Policy Briefings*. <https://bit.ly/3jFocAJ>
- [12] Statistics Korea. (2019). *The social indicator of South Korea*. <https://bit.ly/2XAZINm>
- [13] National Council on Climate and Air Quality (2020). *Report on the current state and solutions for fine particulate air pollution*. <https://bit.ly/3Al1gUw>
- [14] H. M. Byun, Y. J. Park & K. Yun. (2021). An Exploratory Study on the Policy for Facilitating of Health Behaviors Related to Particulate Matter: Using Topic and Semantic Network Analysis of Media Text. *Journal of Korean Academy of Nursing*, 51(1), 68–79. DOI : 10.4040/jkan.20213
- [15] C. G. Choi, & C. M. Kim. (2016). Path Dependency and Social Amplification of Risk in Particulate Matter Air Pollution Management and its Implications. *Journal of The Korean Regional Development Association*, 28(5), 89–107. <https://bit.ly/3bbCz1N>
- [16] Y. Kim, et al. (2015). How Does Media Construct Particulate Matter Risks? : A News Frame and Source Analysis on Particulate Matter Risks. *Korean Journal of Journalism & Communication Studies*, 59(2), 121–154. DOI : 10.20879/kjics.2018.62.4.004
- [17] Y. Kim, et al. (2017). A Study on the Environmental Risk Information Seeking and Processing Model about Particulate Matter: Focusing on the Moderating Effects of China Attribution, Health Symptom Experience, Perceived Information Capacity, and Relevant Channel Beliefs. *Korean Journal of Communication Studies*, 25(2), 5–44. DOI : 10.23875/kca.25.2.1
- [18] J. W. Rhee, & S. H. Kim. (2018). News frames in the coverage of fine-dust disaster—Application of Structural Topic Modeling. *Korean Journal of Journalism & Communication Studies*, 62(4), 125–158. DOI : 10.20879/kjics.2018.62.4.004
- [19] World Health Organization. (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*. Geneva: World Health Organization. <https://bit.ly/3vL0238>
- [20] S. Kyung, & S. H. Jeong. (2017). Adverse health effects of particulate matter. *Journal of the Korean Medical Association*, 60(5), 391–398. DOI : 10.5124/jkma.2017.60.5.391
- [21] G. D. Thurston, et al. (2017). A joint ERS/ATS policy statement: what constitutes an adverse health effect of air pollution? An analytical framework. *European Respiratory Journal*, 49(1). DOI : 10.1183/13993003.00419–2016
- [22] OECD (2016). *The Economic Consequences of Outdoor Air Pollution*. Organization for Economic Cooperation and Development. <https://bit.ly/3menCCn>
- [23] S. E. Ahn, et al. (2017). *Assessment of Human Health Effects of Air-Pollution Using Cohort DB and Estimation of Economic Costs in Korea (III)*. Korea Environment Institute. <https://bit.ly/3EbH9JG>
- [24] Y. Cho, & Y. H. Sonn. (2004). Estimation of the Health Benefits for Improving Air Quality. *Korea Review of Applied Economics*, 6(1), 133–150. G704–001413.2004.6.1.004
- [25] J. W. Min. (2019). A Survey of Citizens' Perception of Fine Particulate Matter. *HRI Weekly Economic Review*, 833, 1–16. <https://bit.ly/3CfU0KI>
- [26] S. K. Cho, J. I. Chang, & J. I. Kim. (2006). Valuing the Health Effects on Air Quality Improvement—Using Conjoint Analysis. *Environmental and Resource Economics Review*, 15(5), 859–884. <https://bit.ly/3vJSYnz>
- [27] S. J. Yoon. (2009). Finding governance to prevent and mitigate social conflicts surrounding climate change response. *Journal of Governance Studies*, 4(2), 125–160. <https://bit.ly/3nABVAS>
- [28] W. Chung. (2021). The Discourse Analyses of the Online News Articles and Replies on the Fine Dust Risk Issue : Based on the Issue Life Cycle Model. *Korean Association for Advertising and Public Relations*, 23(2), 140–183. <https://bit.ly/3GpGshS>
- [29] M. Kam & M. Song. (2012). A Study on Differences of

Contents and Tones of Arguments among Newspapers Using Text Mining Analysis. *Journal of Intelligence and Information Systems*, 18(3), 53-77.
DOI : 10.13088/jiis.2012.18.3.053

- [30] A. Kim, et al. (2018). An Analysis of Effects of Emergency Fine Dust Reduction Measures and National Petition Using Regression Analysis and Text Mining. *KIPS Transactions on Software and Data Engineering*, 7(11), 427-434.
DOI : 10.3745 /KTSDE.2018.7.11.427
- [31] S. Yoon & M. Kim. (2020). Topic Modeling on Fine Dust Issues Using LDA Analysis. *Journal of Energy Engineering*, 29(2), 23-29.
DOI : 10.5855 /ENERGY.2020.29.2.023
- [32] J. H. Gu, et al. (2019). Policy Trends and Environmental Studies through Big Data. Sungkyunkwan University. bit.ly/3GpGshS
- [33] BIG KINDS. Korea Press Foundation(KPF). <https://www.bigkinds.or.kr>
- [34] AIRKOREA. Republic of Korea: Ministry of Environment. <https://www.airkorea.or.kr>
- [35] D. Jin, et al. (2018). *A Study on the Utilization of Text Mining for Climate Environmental Issues Analysis*. Korea Environment Institute. <https://bit.ly/3nHr4Xw>
- [36] T. Hardeniya & D. A. Borikar. (2016). Dictionary based approach to sentiment analysis—a review. *International Journal of Advanced Engineering, Management and Science*, 2(5), 317-322.
<https://bit.ly/3Eq1hJ0>
- [37] S. M. Park, et al. (2018). KNU Korean Sentiment Lexicon: Bi-LSTM-based Method for Building a Korean Sentiment Lexicon. *Journal of Intelligence and Information Systems*, 24(4), 219-240.
DOI : 10.13088/jiis.2018.24.4.219
- [38] S. Choi. (2018). EasySenti: Easy Sentiment Analysis. github.com/SukjaeChoi/easySenti
- [39] National Information Society Agency. (2017). *The Issue of Particulate Matter and the Application of Big Data*. NEAR & Future INSIGHT. <https://bit.ly/3zn8aqK>

양 지 연(Ji-Yeon Yang)

[충원]



- 2010년 7월 : Univ. of Illinois Urbana-Champaign 통계학 박사
- 2010년 7월 ~ 2011년 6월 : Claremont McKenna College 방문조교수
- 2011년 7월 ~ 2014년 2월 : MD Anderson Cancer Center 연구원
- 2014년 3월 ~ 현재 : 금오공과대학교

- 응용수학과 조교수/부교수
- 관심분야 : Bayesian analysis, big data analytics and computational statistics.
- E-Mail : jyang@kumoh.ac.kr